confluent.io/ksql

Apache Kafka and KSQL in Action : Let's Build a Streaming Data Pipeline!

JZ2018

@rmoff robin@confluent.io

https://cnfl.io/ksql-workshop-prereqs

- Make sure you allocate Docker >=8GB memory

```
docker system info | grep Memory
```

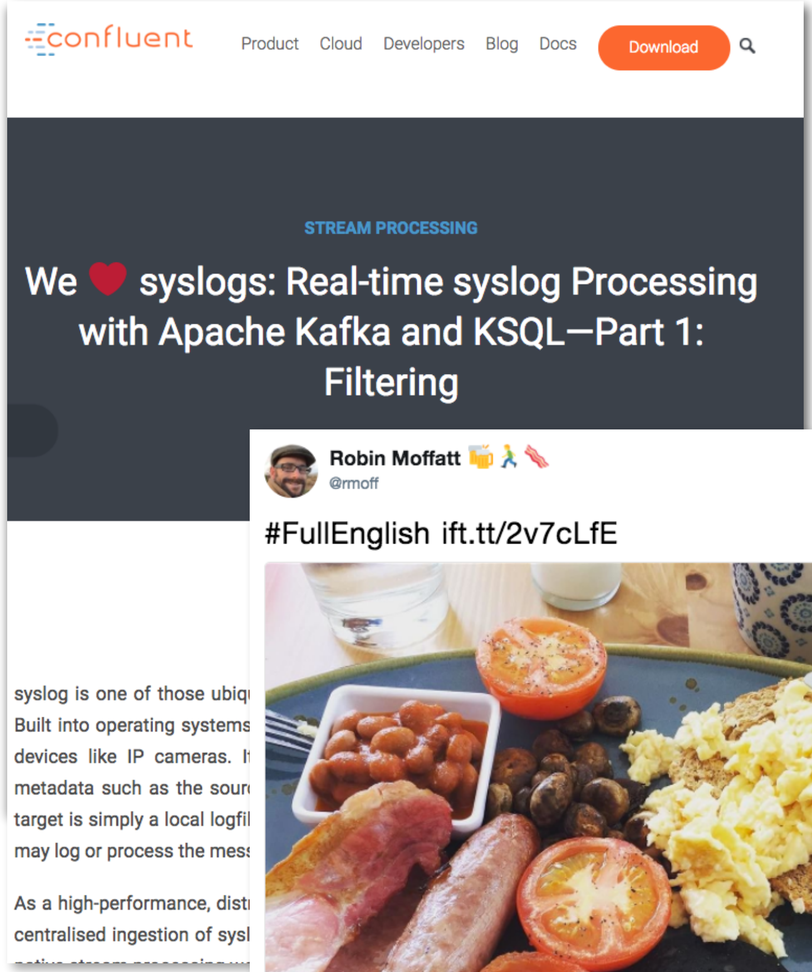
- Clone the repo
  - Should default to branch 5.0.0-post
- Pull the git images as instructed in the doc

https://cnfl.io/ksql-workshop

3: Start up the Stack

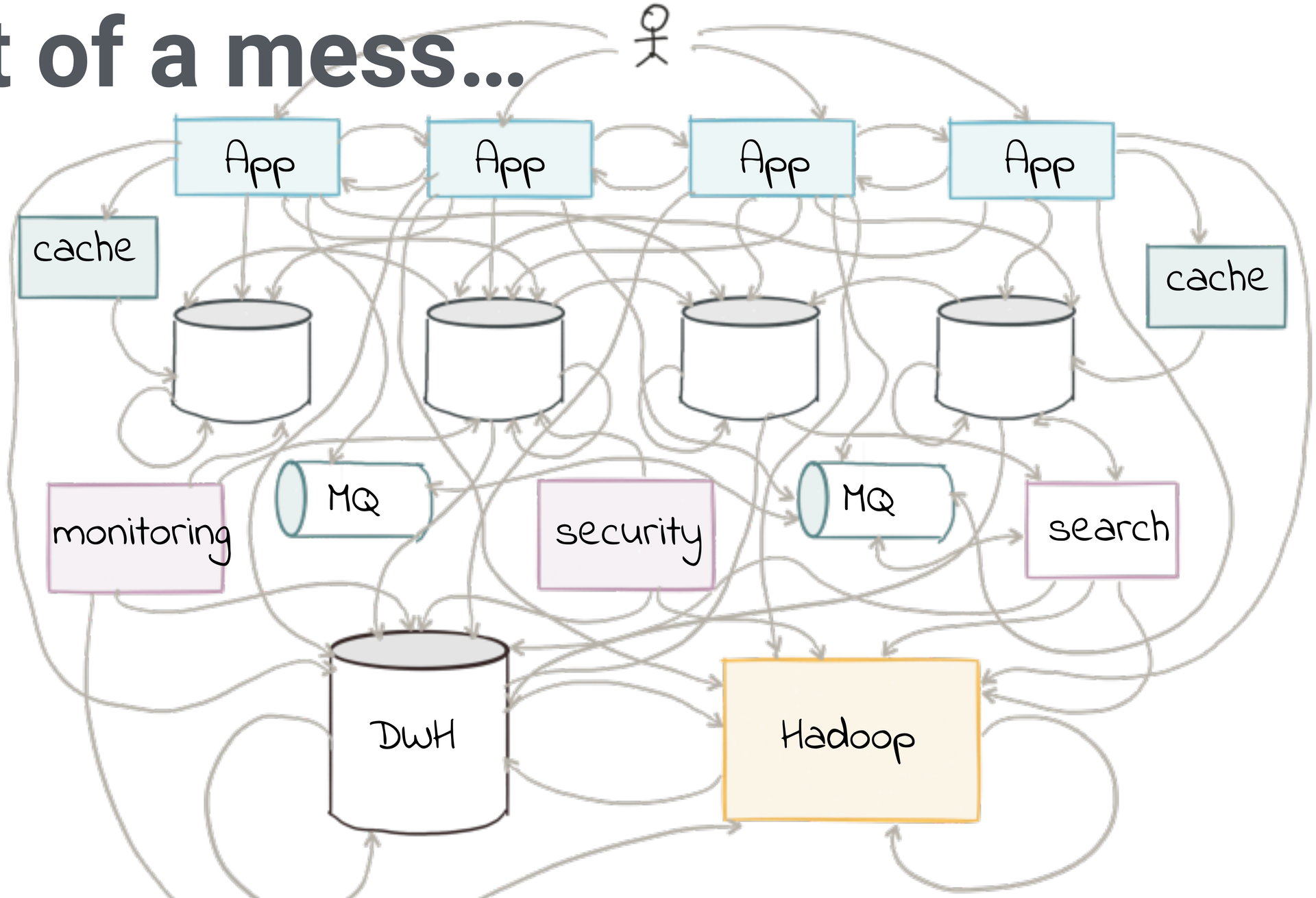
$ whoami

- Developer Advocate @ Confluent
- Working in data & analytics since 2001
- Oracle ACE Director & Dev Champion
- Blogging : http://rmoff.net & http://cnfl.io/rmoff
- Twitter: `@rmoff`
  - Geek stuff
  - Beer & Fried Breakfasts

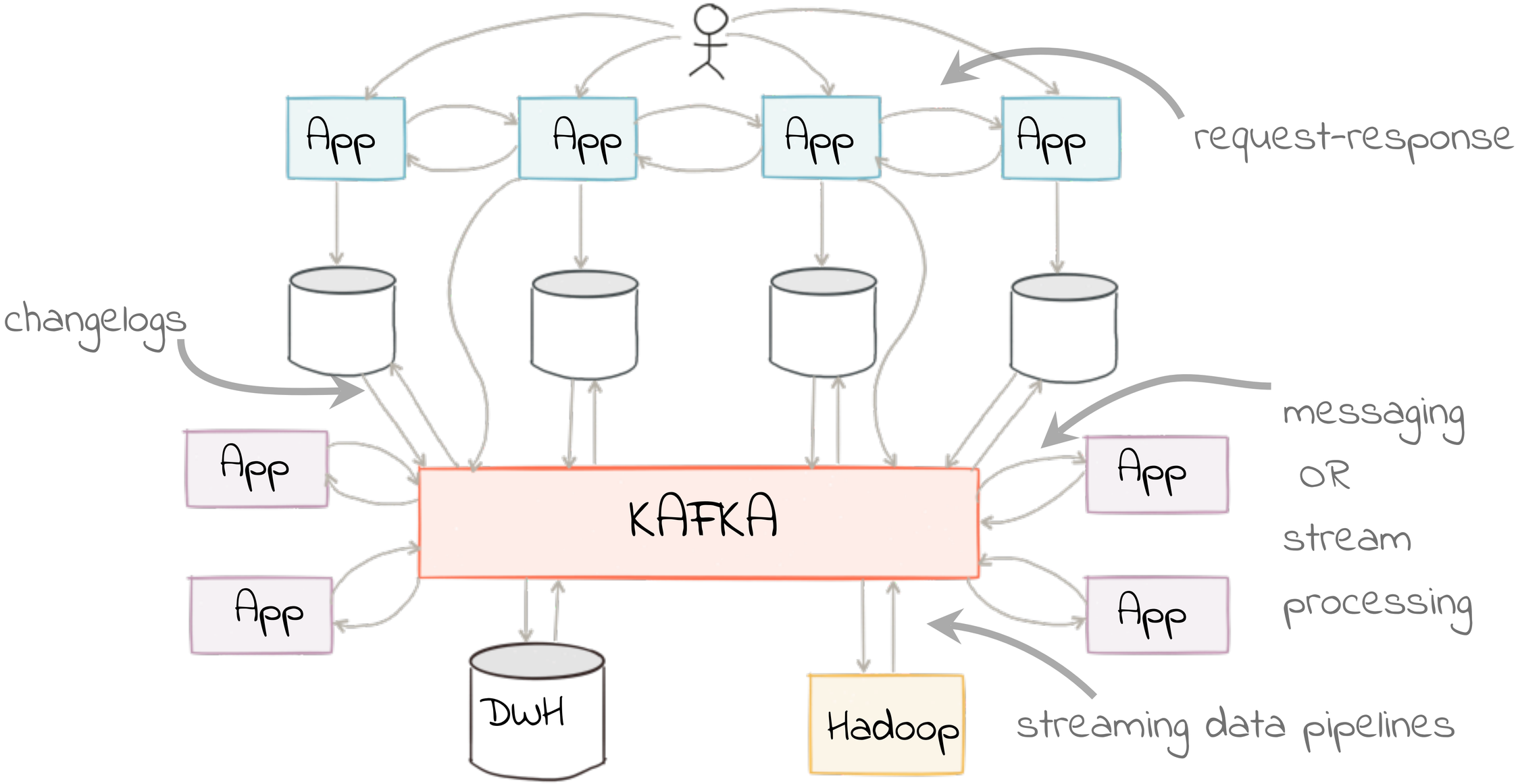


https://speakerdeck.com/rmoff/

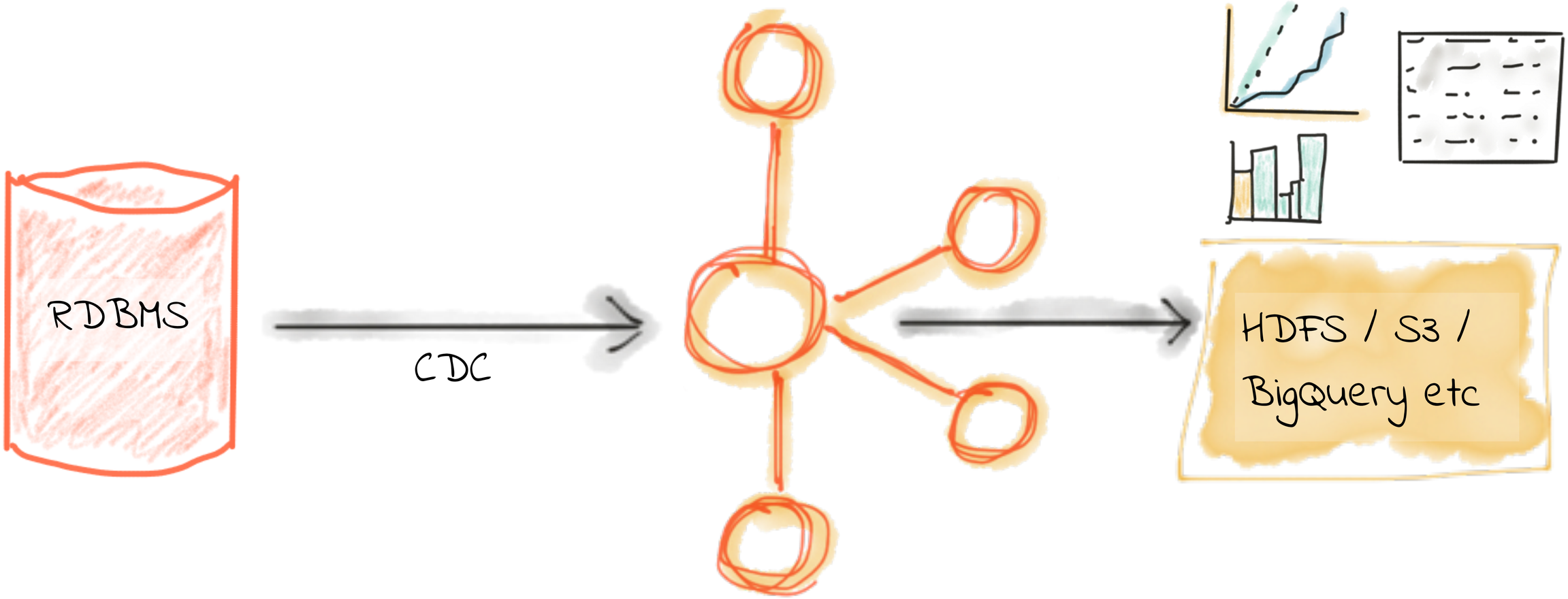
A bit of a mess…



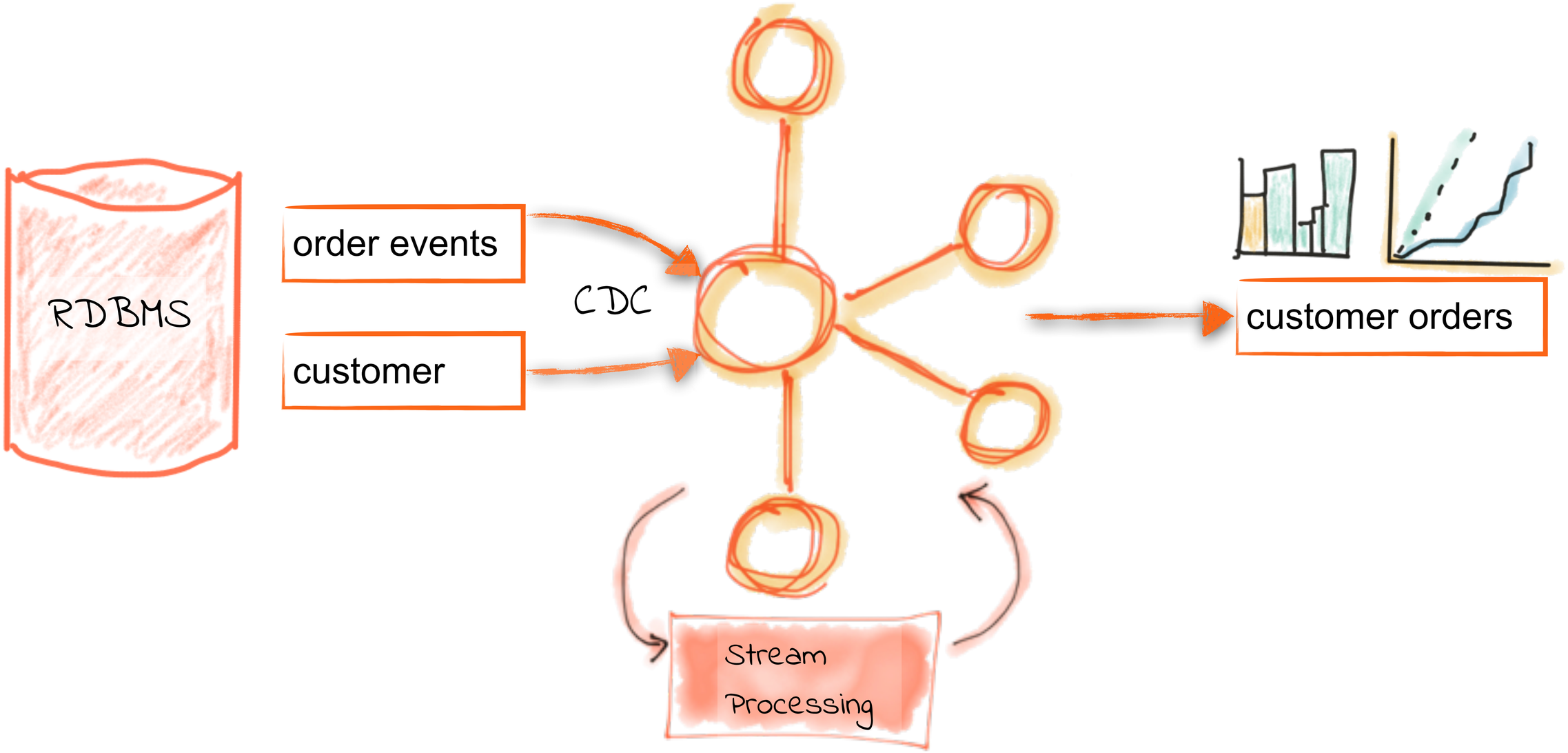
Kafka is a Streaming Platform



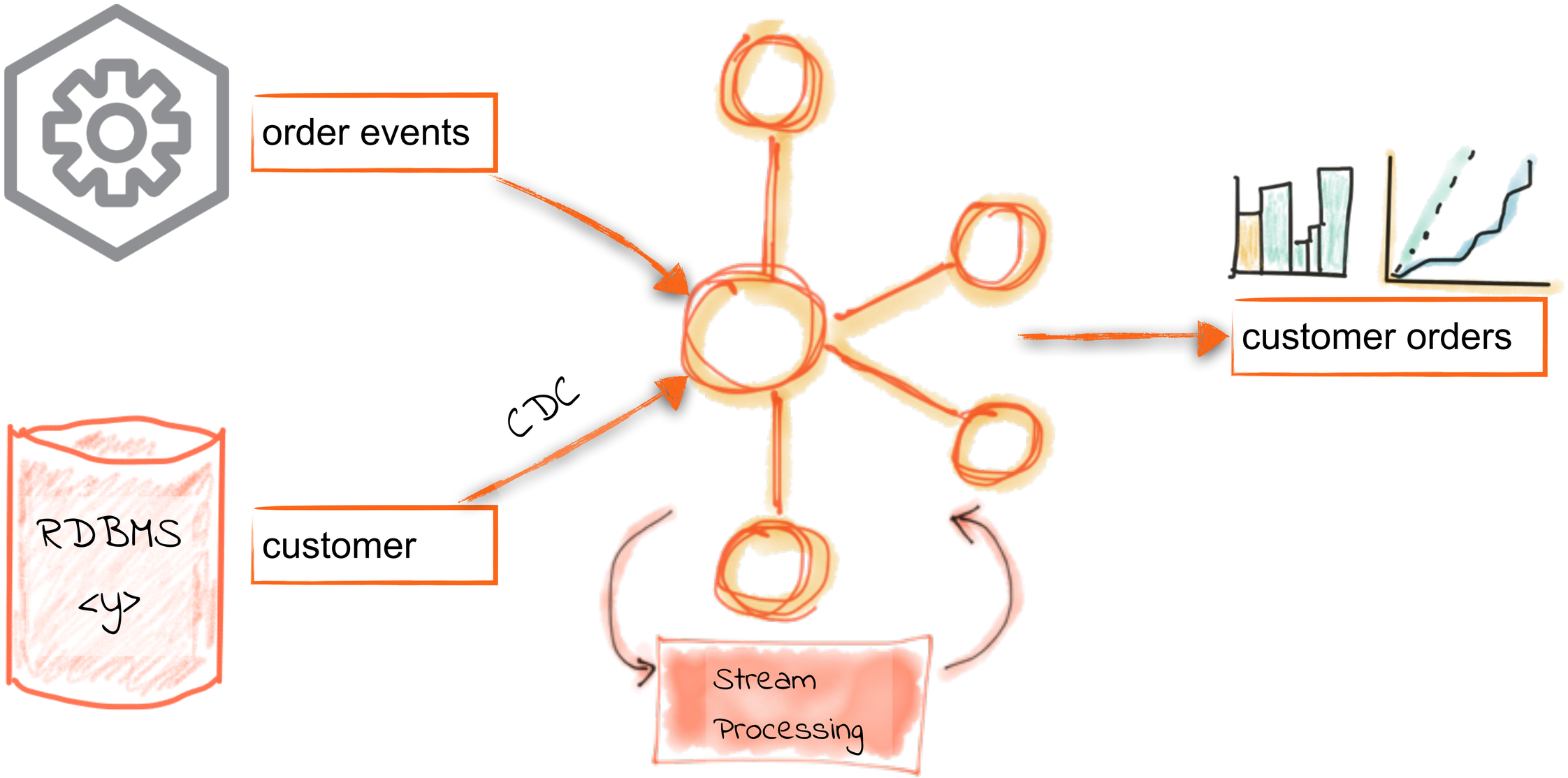
Analytics - Database Offload



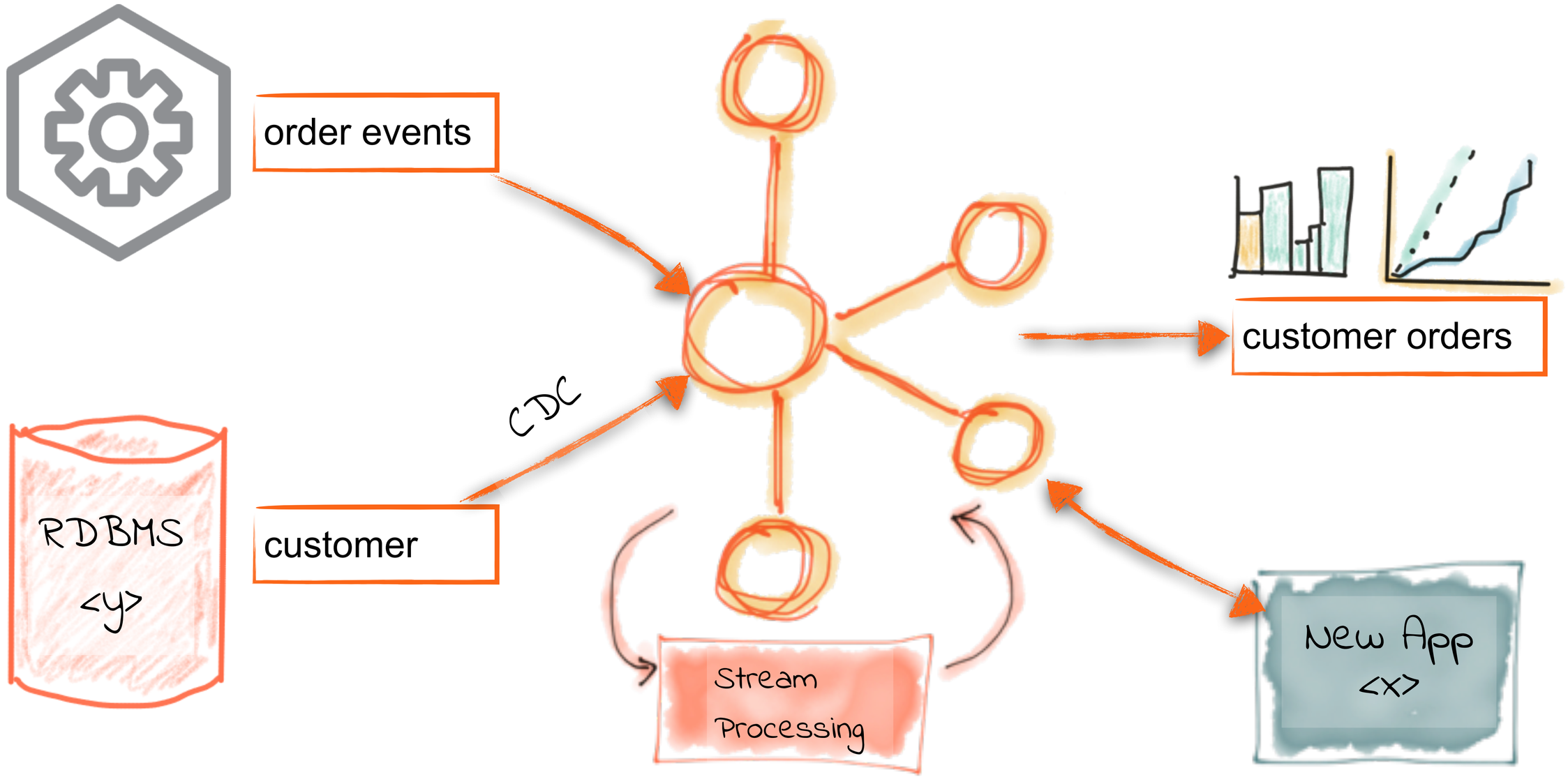
Stream Processing with Apache Kafka and KSQL



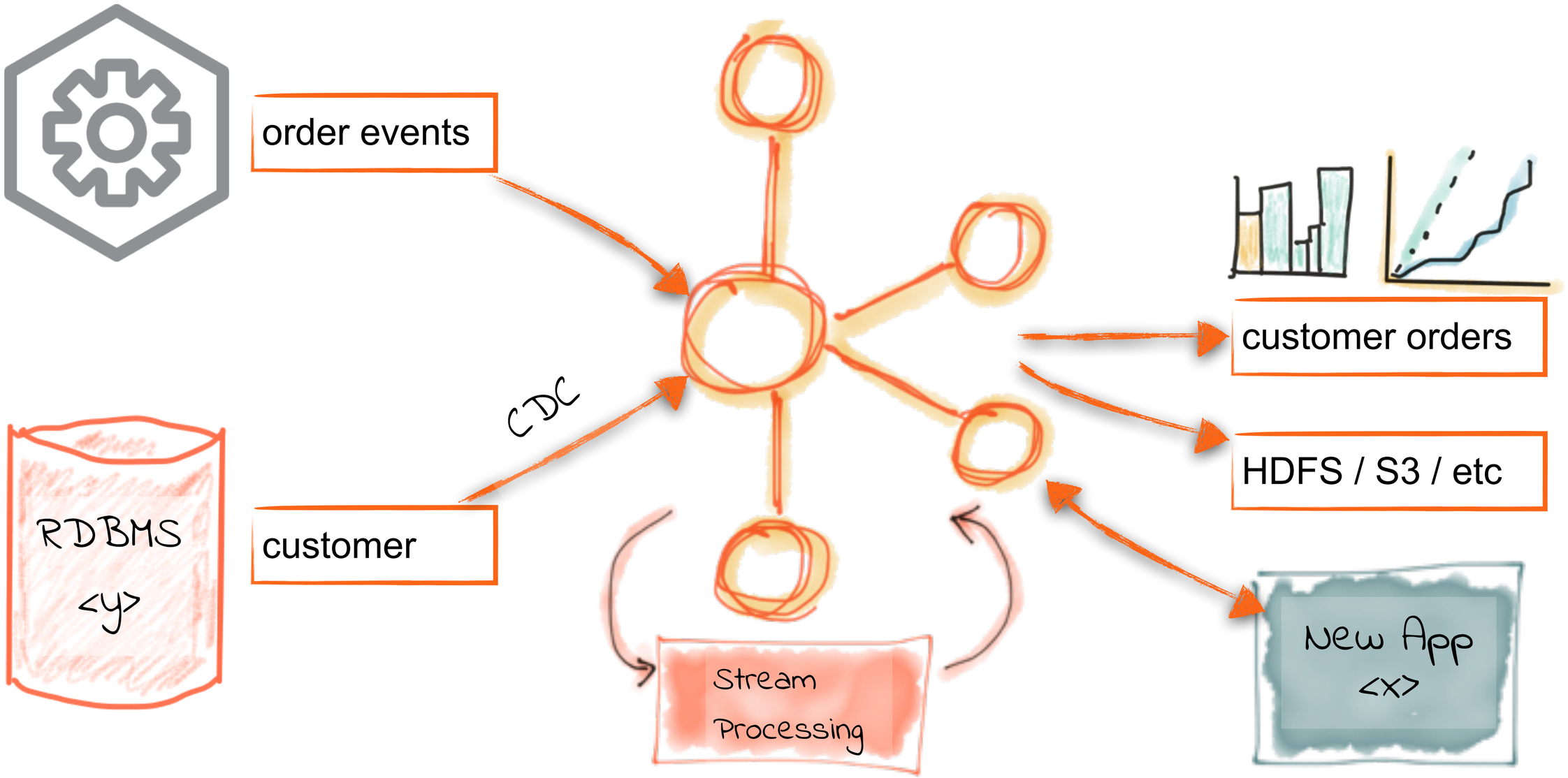
Real-time Event Stream Enrichment



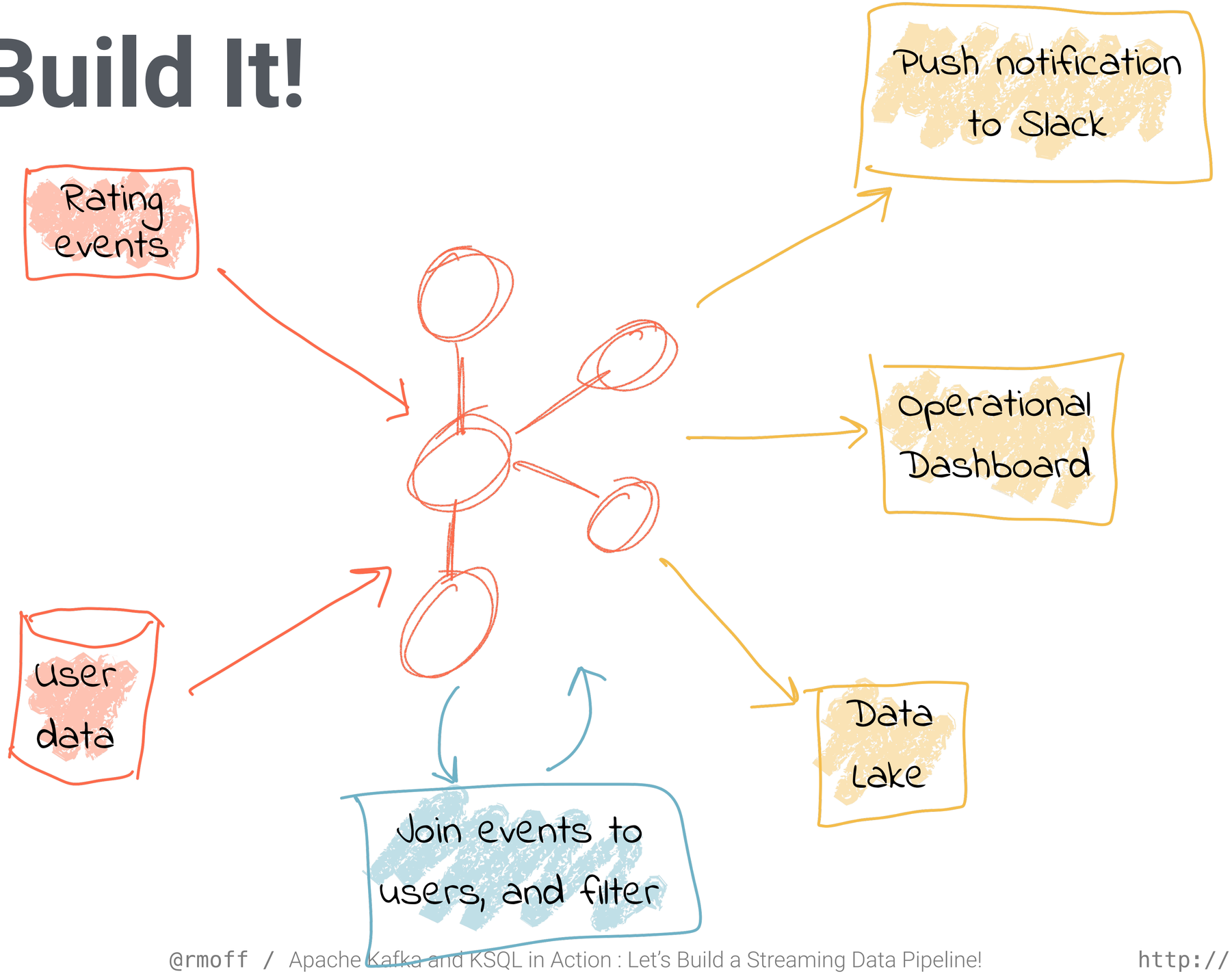
Transform Once, Use Many



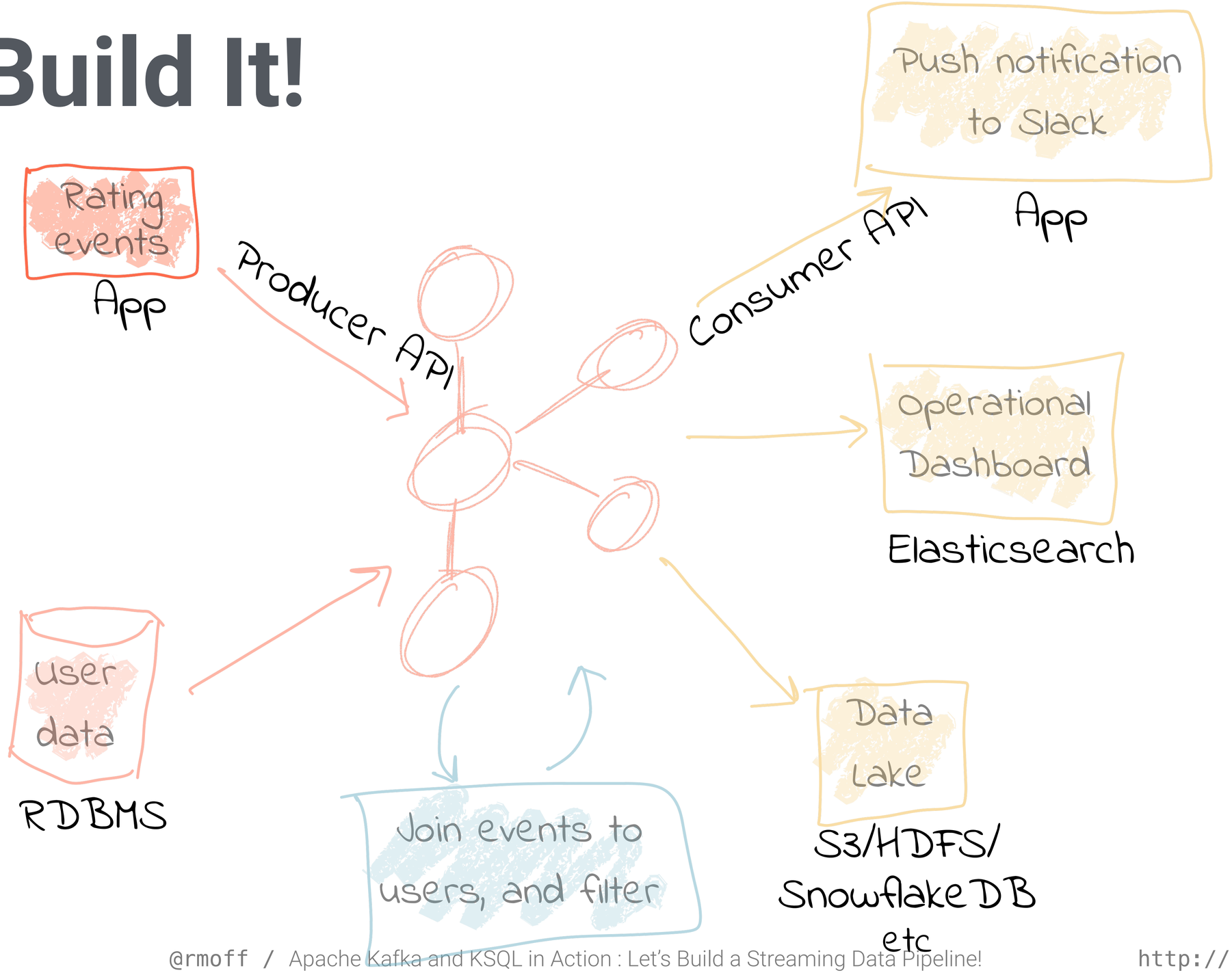
Transform Once, Use Many



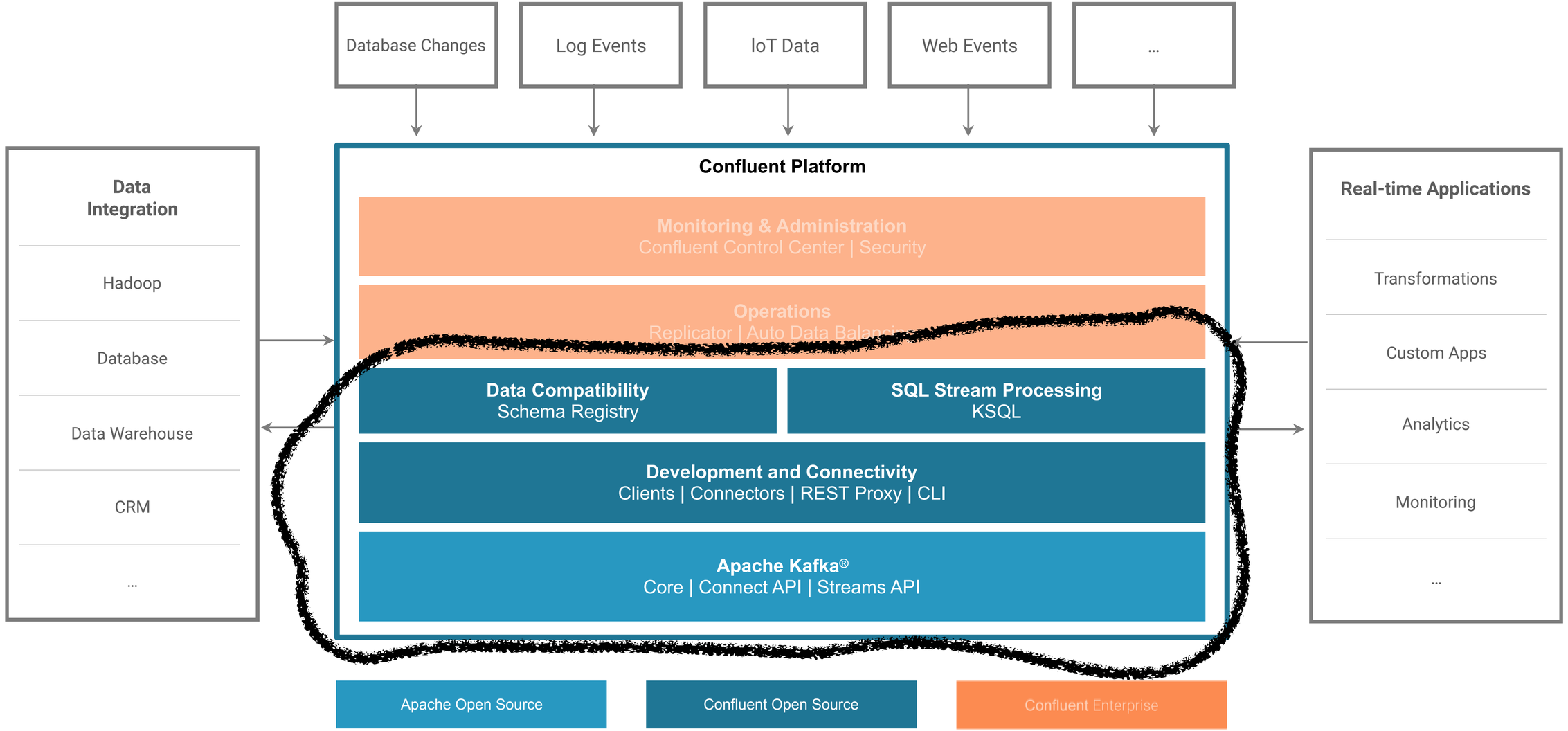
Let's Build It!



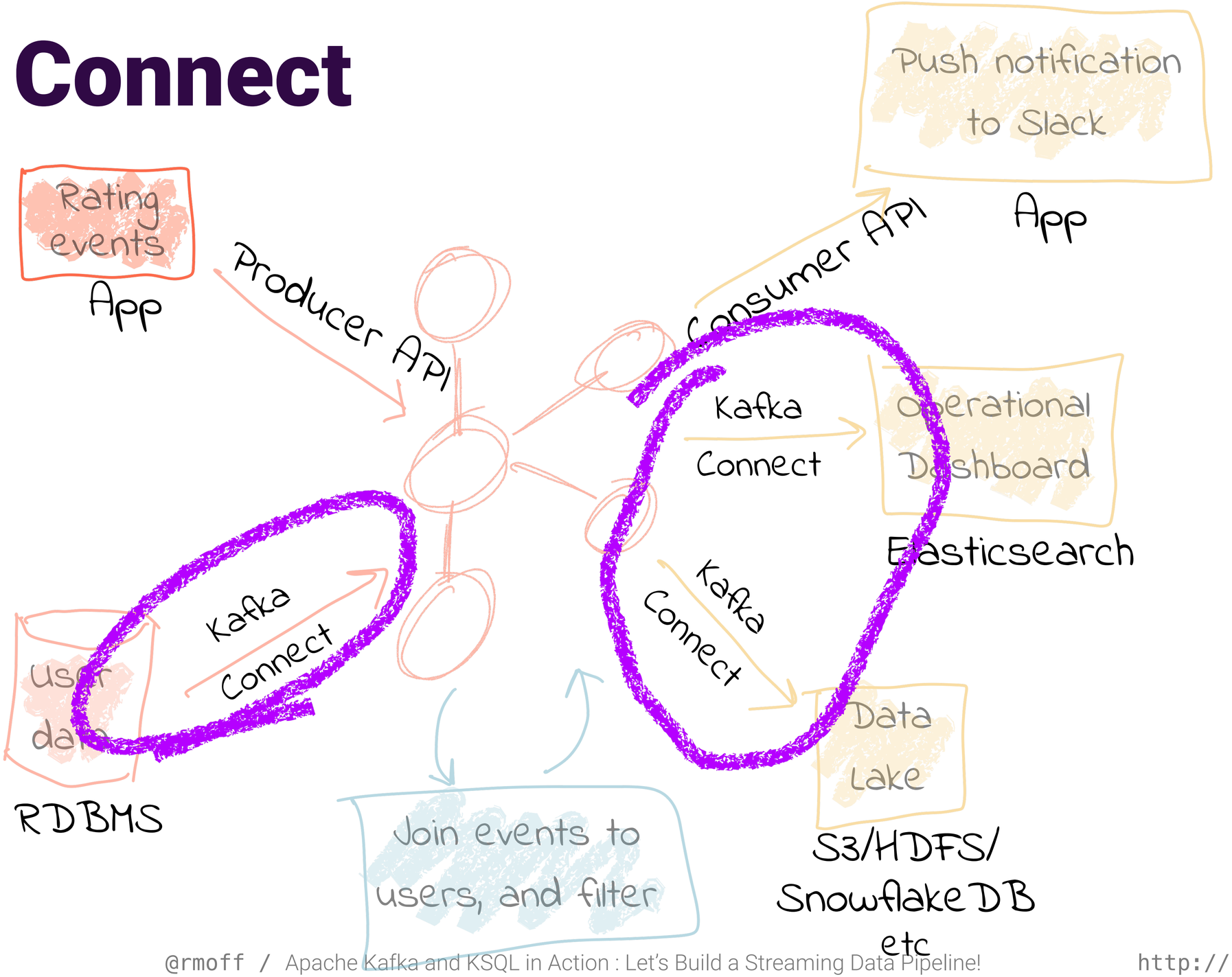
Let's Build It!



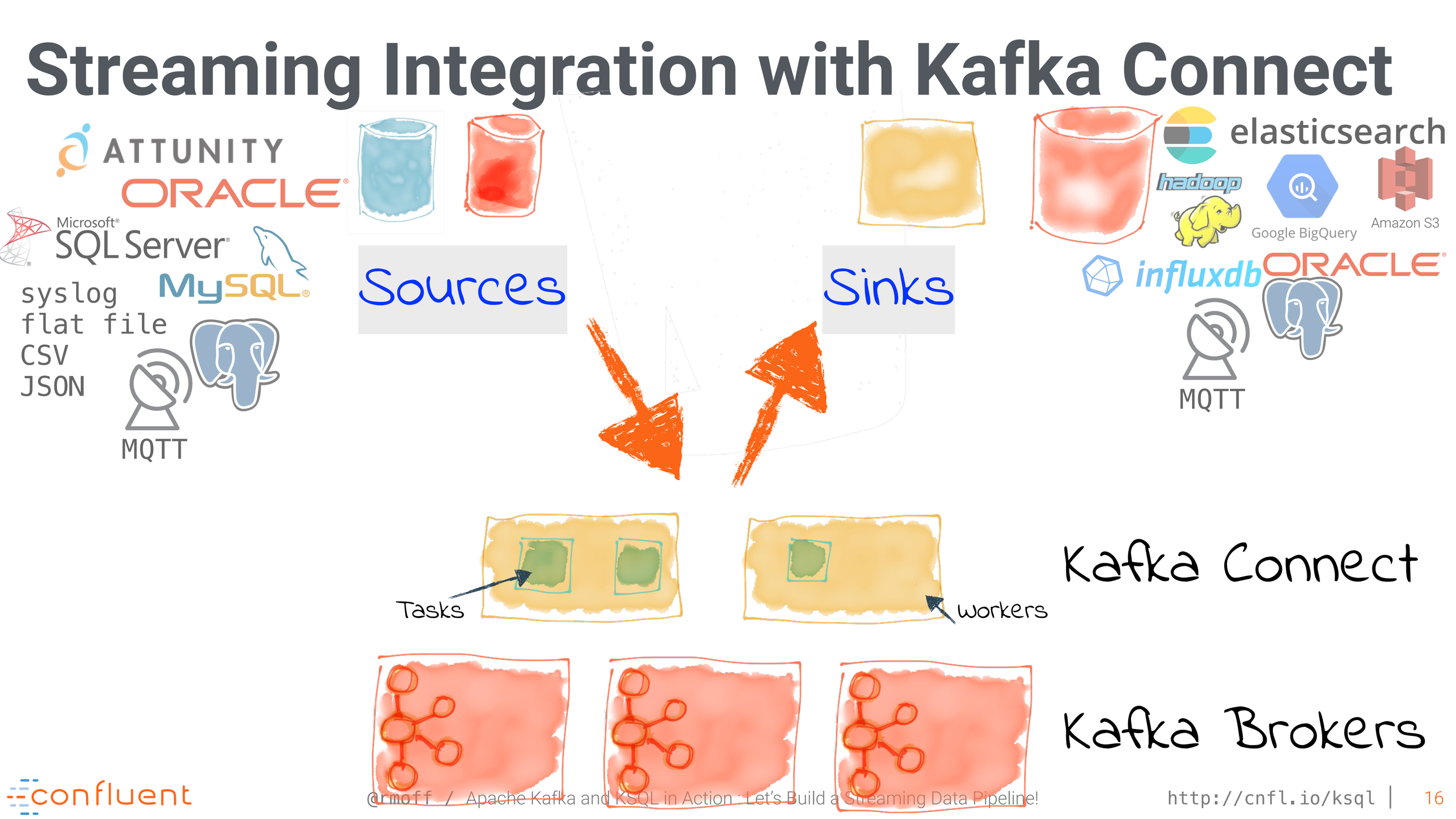
Confluent Open Source : Apache Kafka with a bunch of cool stuff! For free!



Kafka Connect



Streaming Integration with Kafka Connect



Kafka Connect

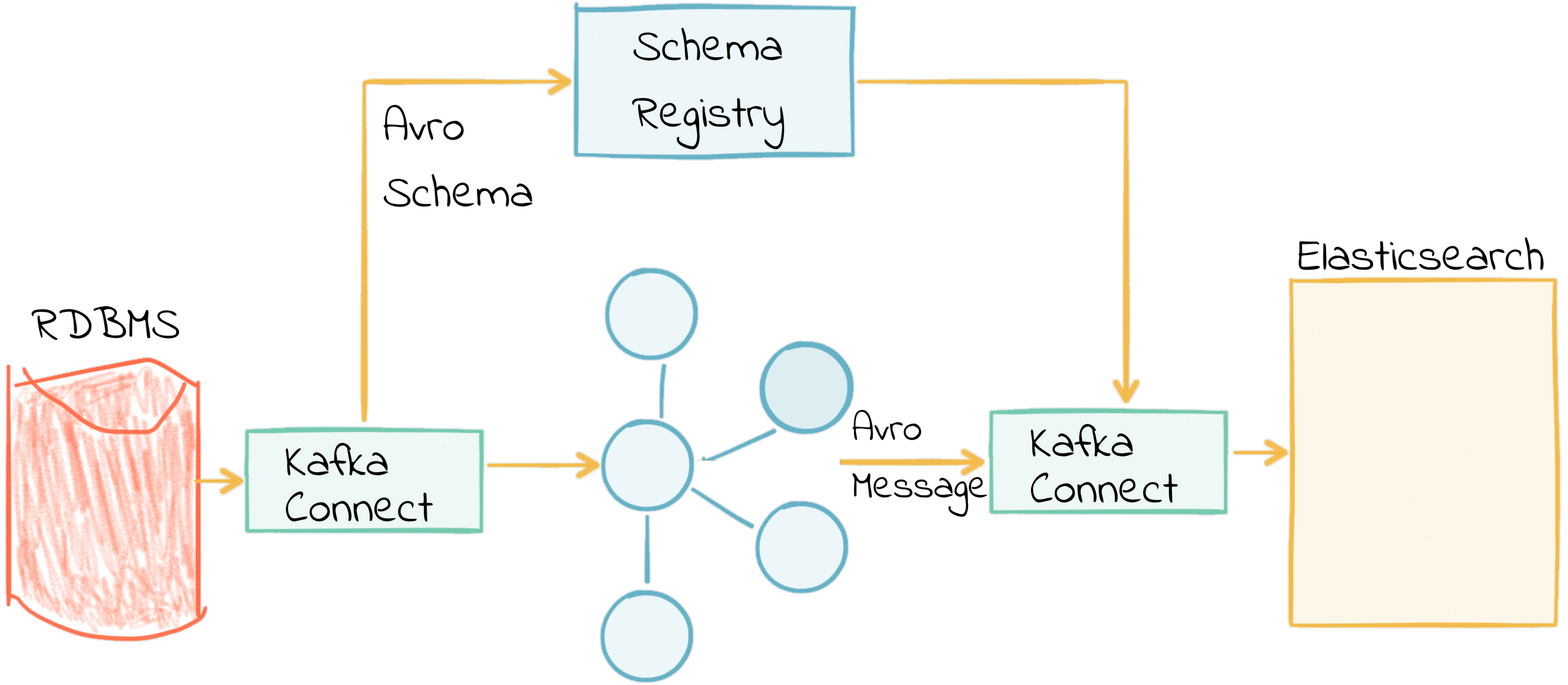
Reliable and scalable integration of Kafka with other systems – no coding required.

- ✓ Centralized management and configuration
- ✓ Support for hundreds of technologies including RDBMS, Elasticsearch, HDFS, S3
- ✓ Supports CDC ingest of events from RDBMS
- ✓ Preserves data schema
- ✓ Fault tolerant and automatically load balanced
- ✓ Extensible API
- ✓ Single Message Transforms
- ✓ Part of Apache Kafka, included in Confluent Open Source

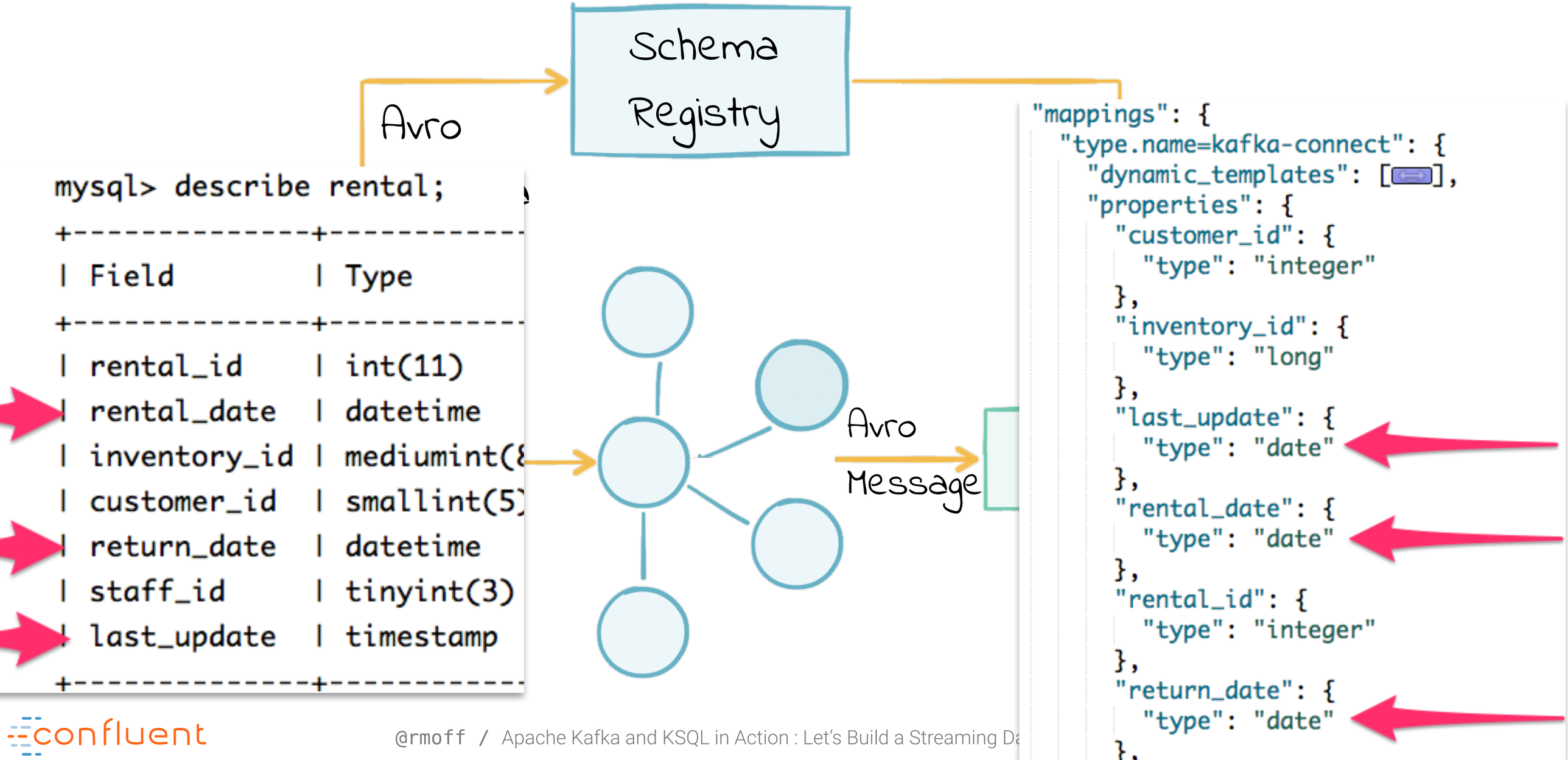
```
{
"connector.class": "io.confluent.connect.jdbc.JdbcSourceConnector",
"connection.url": "jdbc:mysql://localhost:3306/demo?user=rmoff&password=foo",
"table.whitelist": "sales,orders,customers"
}
```

https://docs.confluent.io/current/connect/

Kafka Connect + Schema Registry = WIN

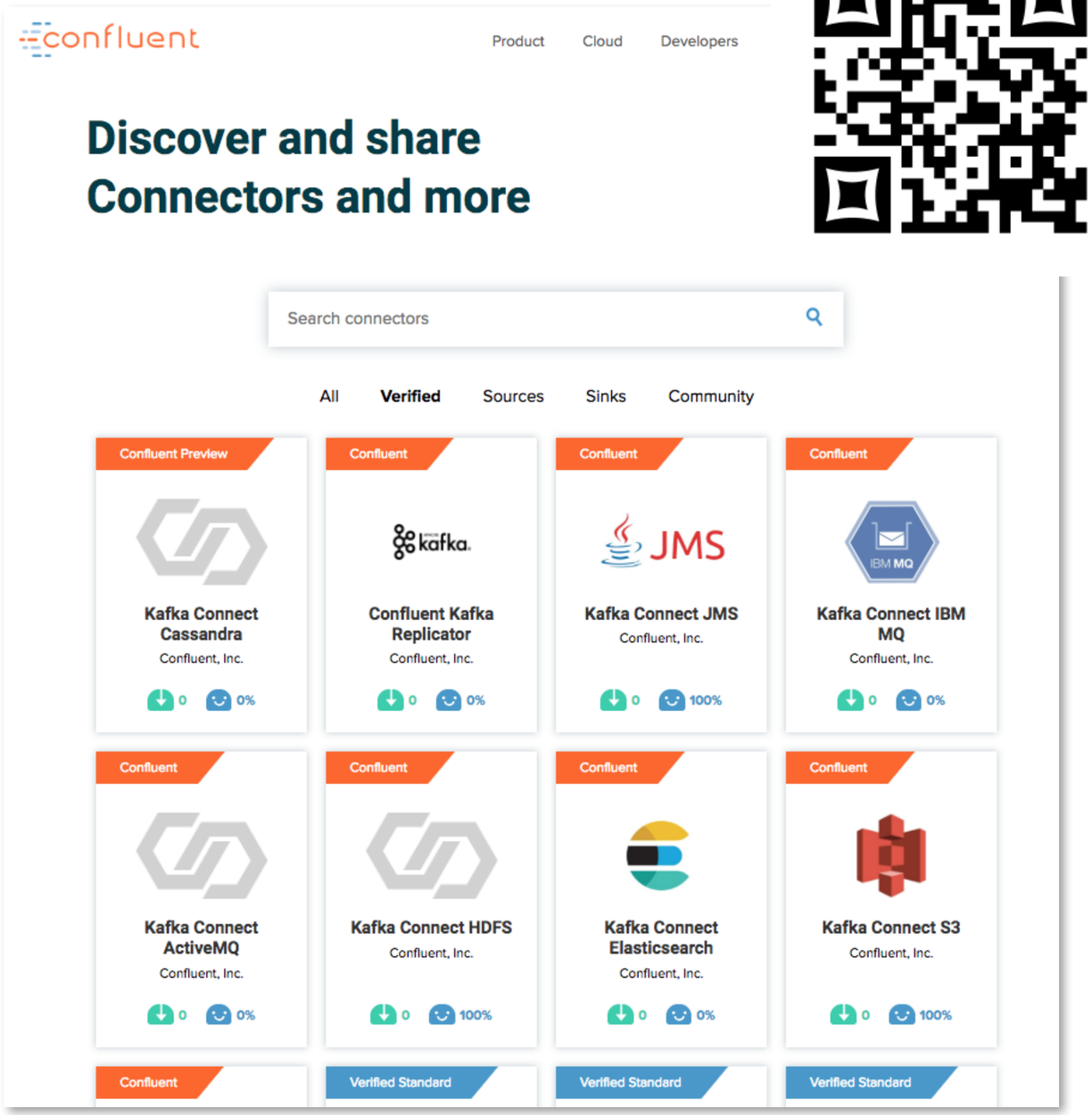


Kafka Connect + Schema Registry = WIN



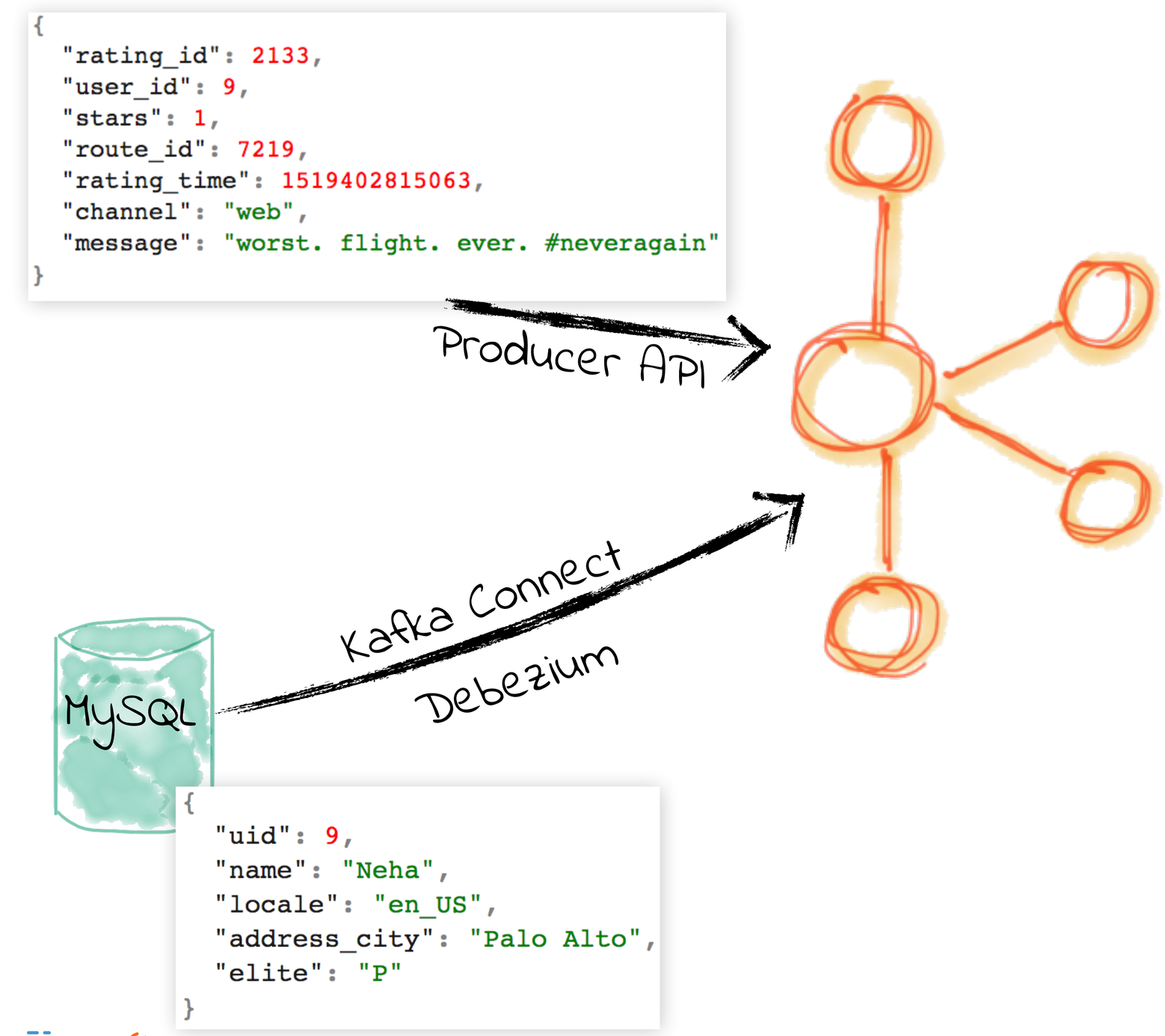
Confluent Hub

- One-stop place to discover and download :
  - Connectors
  - Transformations
  - Converters



hub.confluent.io

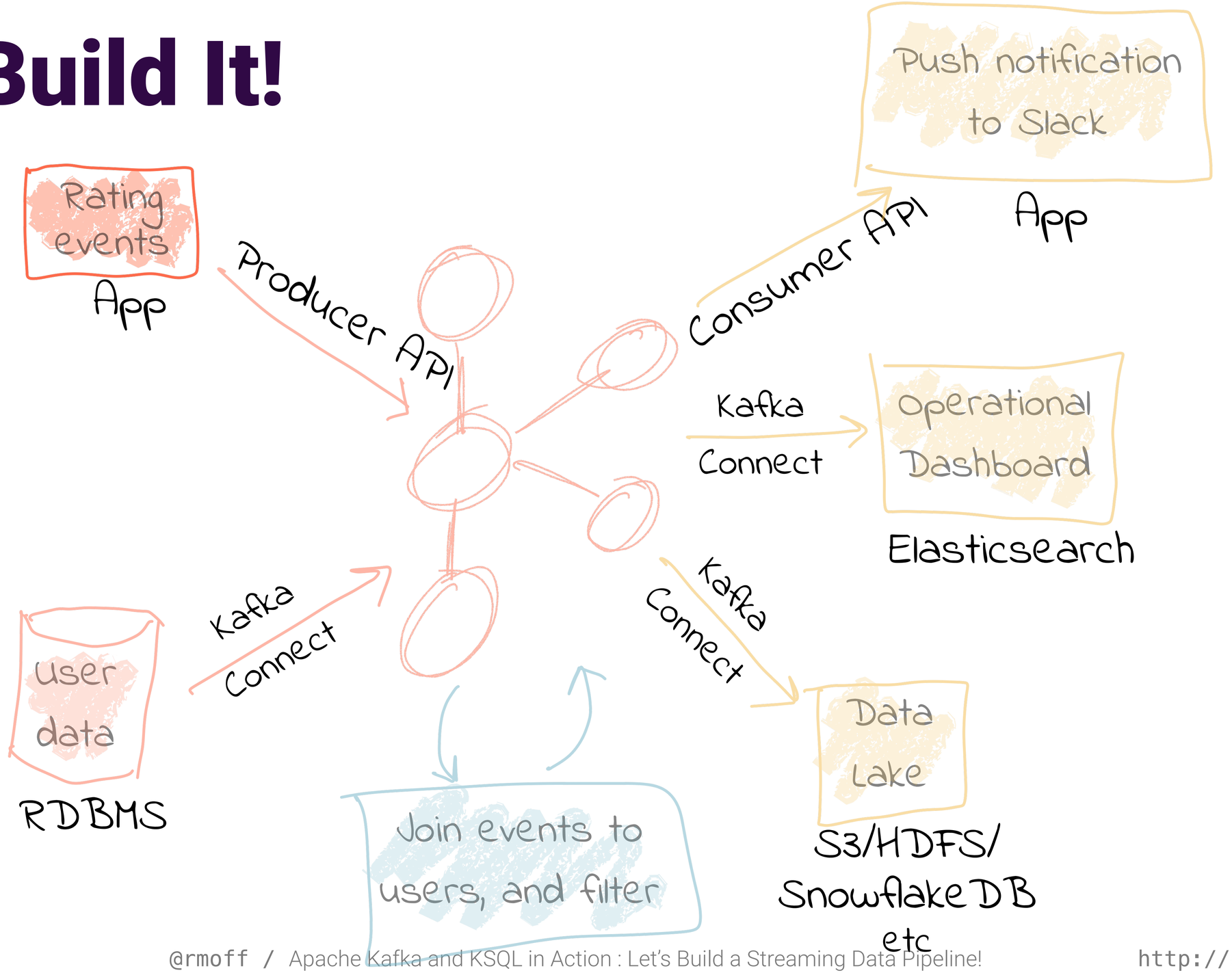
Demo Time!


{
"rating_id": 2133,
"user_id": 9,
"stars": 1,
"route_id": 7219,
"rating_time": 1519402815063,
"channel": "web",
"message": "worst. flight. ever. #neveragain"
}
Producer API
Kafka Connect
Debezium
MySQL
{
"uid": 9,
"name": "Neha",
"locale": "en_US",
"address_city": "Palo Alto",
"elite": "P"
}

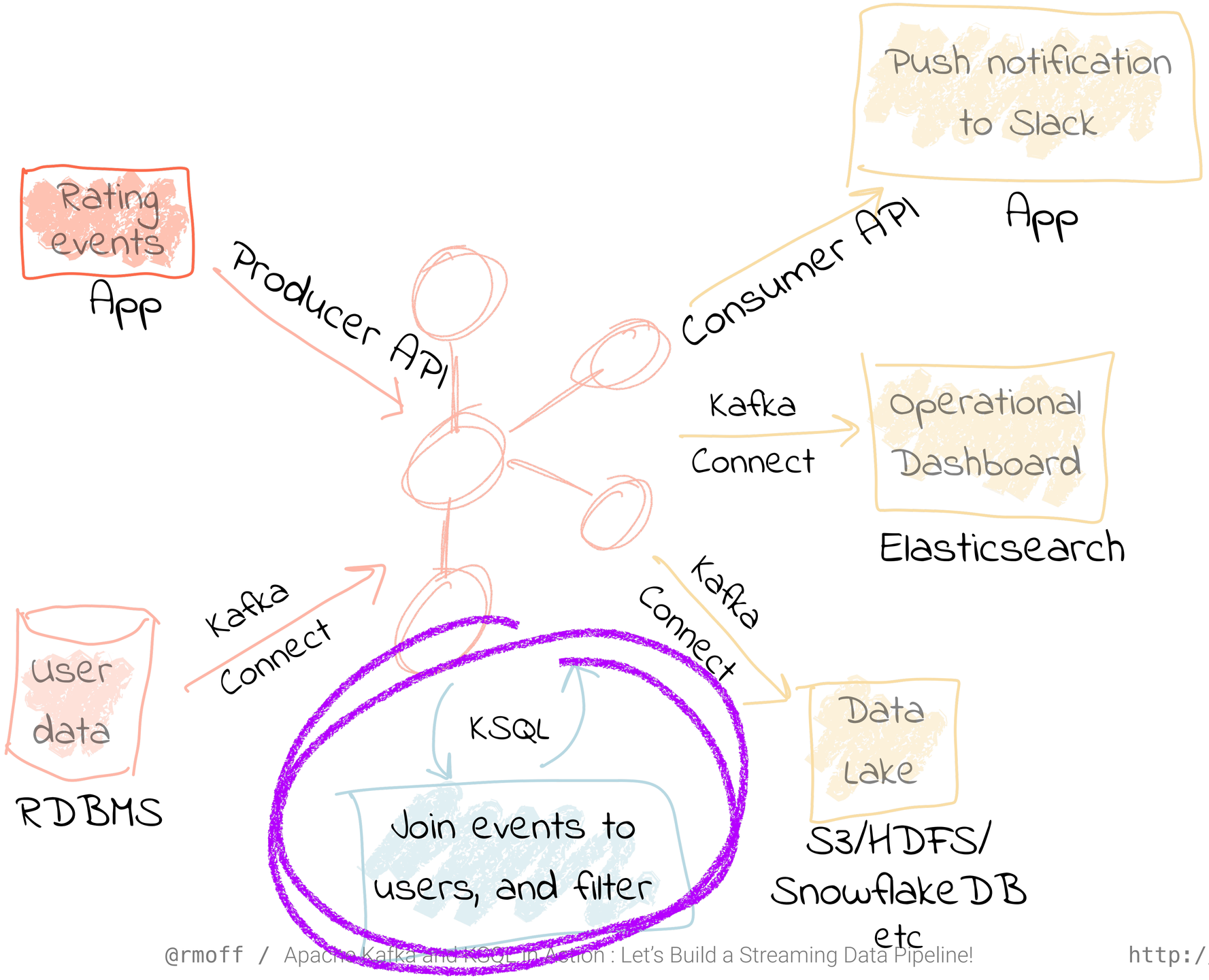
https://cnfl.io/ksql-workshop

4 & 5: Setup & Inspect source data

Let's Build It!



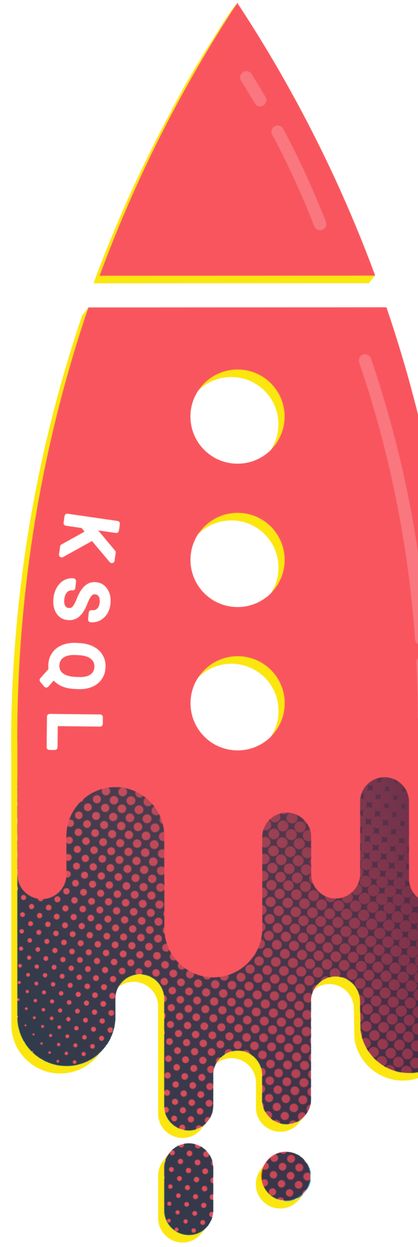
KSQL


Push notification to Slack
App
Rating events
App
Producer API
Consumer API
Kafka Connect
Operational Dashboard
Elasticsearch
Kafka Connect
Kafka Connect
user data
RDBMS
KSQL
Join events to users, and filter
Data Lake
S3/HDFS/ SnowflakeDB etc

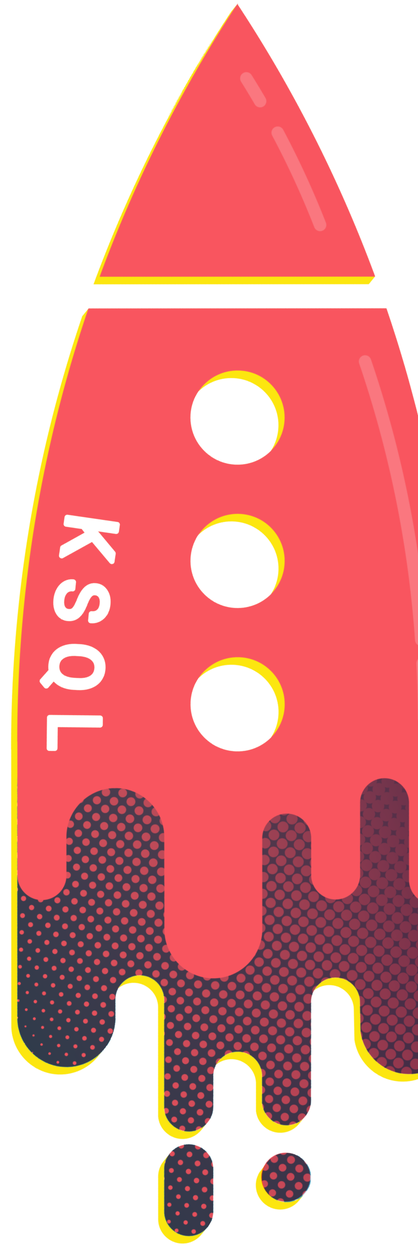
KSQL

is a

Declarative Stream Processing Language



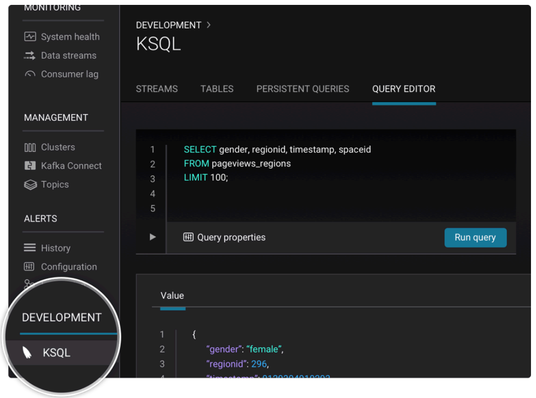
KSQL
is the
Streaming
SQL Engine
for
Apache Kafka



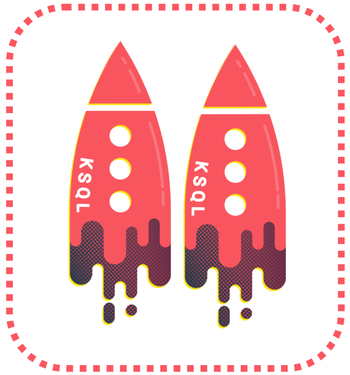
KSQL in Development and Production

Interactive KSQL
for development and testing

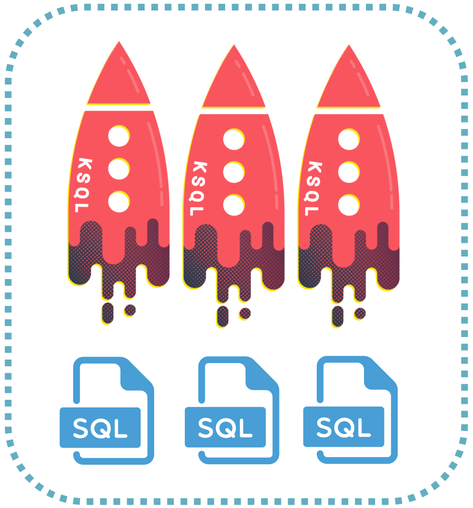
Headless KSQL
for Production



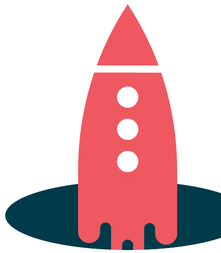
REST



Desired KSQL queries have been identified

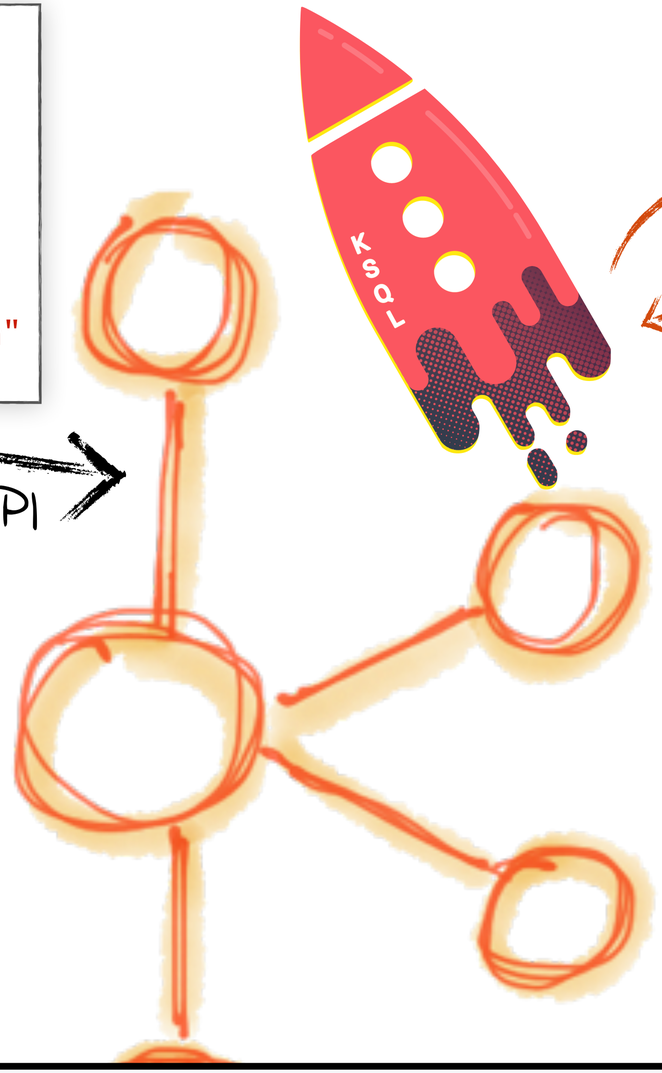


"Hmm, let me try out this idea..."



```
{
 "rating_id": 5313,
 "user_id": 3,
 "stars": 4,
 "route_id": 6975,
 "rating_time": 1519304105213,
 "channel": "web",
 "message": "worst. flight. ever. #neveragain"
}
```

Producer API



Filter all ratings where STARS<3

POOR_RATINGS

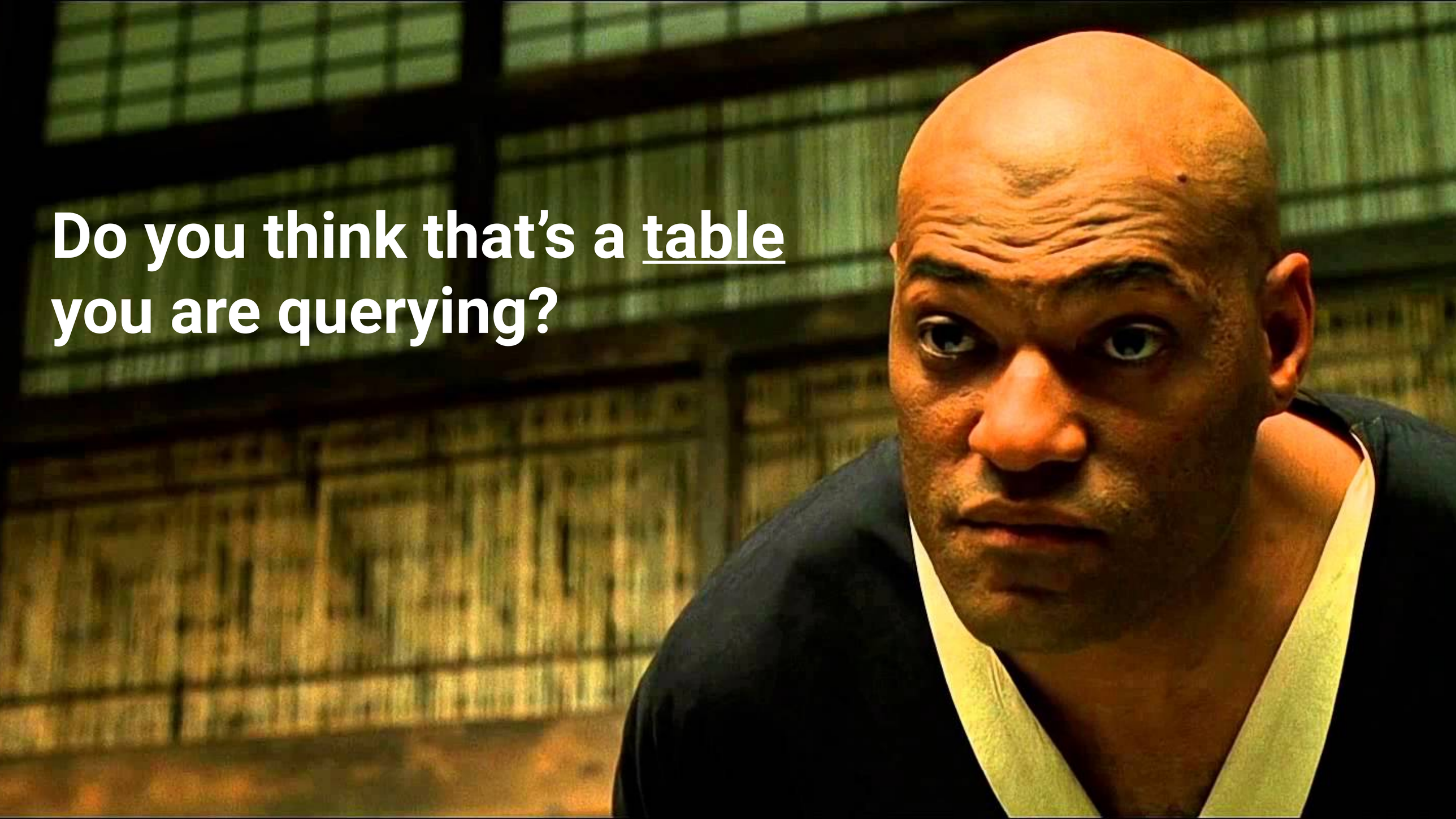
```
CREATE STREAM POOR_RATINGS AS
SELECT * FROM ratings WHERE STARS <3
```

https://cnfl.io/ksql-workshop

6: KSQL CLI

7: Querying the Ratings topic

8. Populating a Kafka topic with KSQL


Do you think that’s a table
you are querying?

The Table Stream Duality

Stream

Account ID	Amount
12345	+ €50
12345	+ €25
12345	-€60

Time

Table

Account ID	Balance
12345	€50

Account ID	Balance
12345	€75

Account ID	Balance
12345	€15

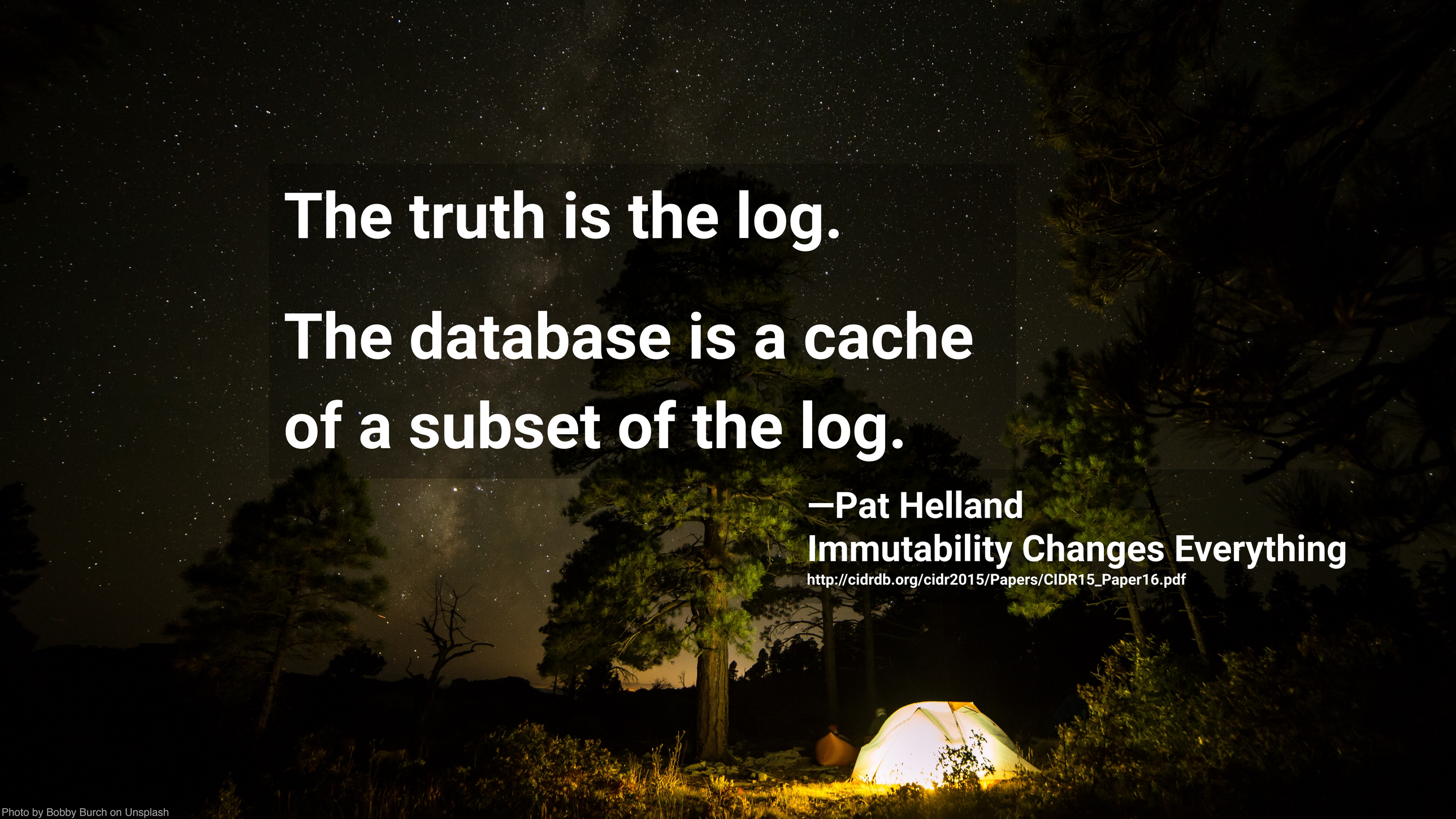
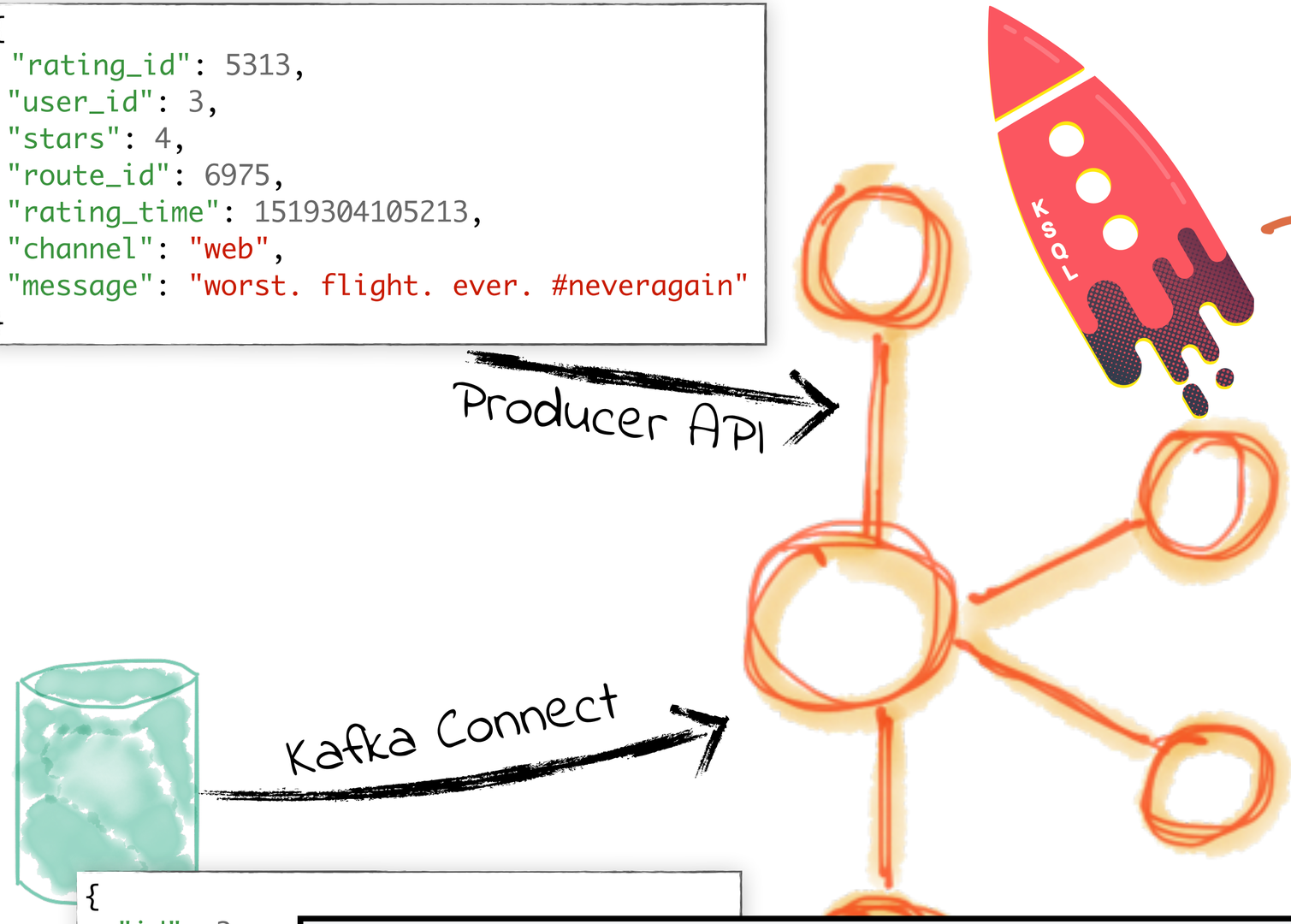


Photo by Bobby Burch on Unsplash

```
{
 "rating_id": 5313,
 "user_id": 3,
 "stars": 4,
 "route_id": 6975,
 "rating_time": 1519304105213,
 "channel": "web",
 "message": "worst. flight. ever. #neveragain"
}
```


Producer API



Join each rating to customer data

RATINGS_WITH_CUSTOMER_DATA

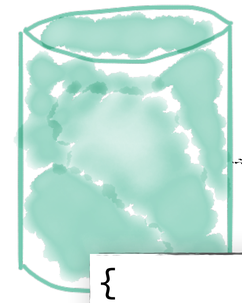
Kafka Connect


```
{
  "id": 3,
  "first_name
  "last_name"
  "email": "m
  "gender": "
  "club_statu
  "comments":
}
```

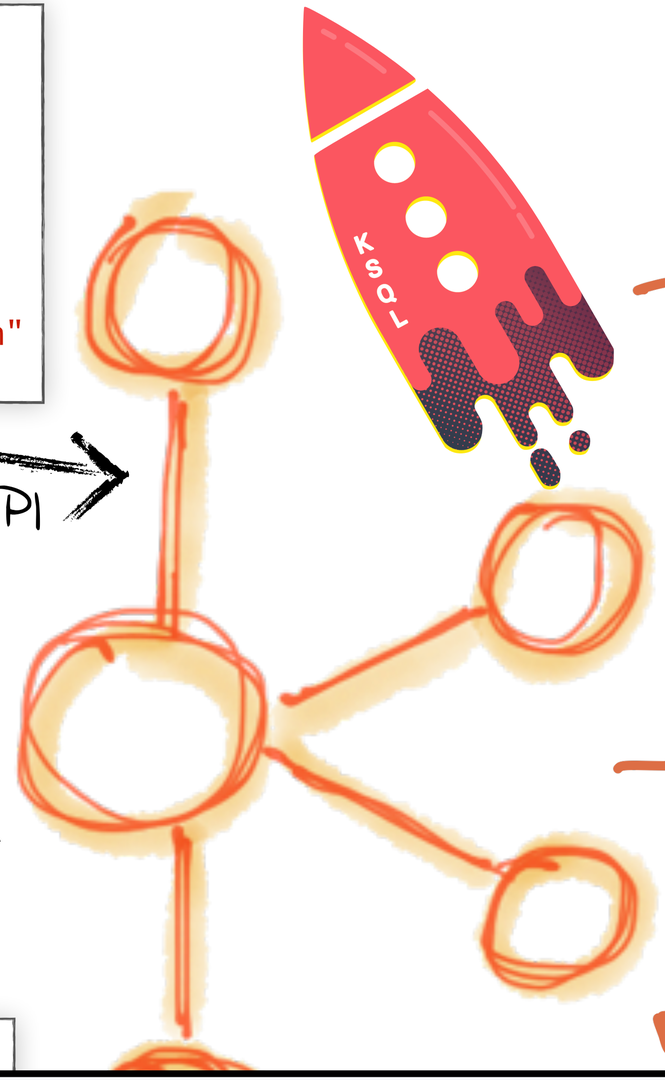

```
CREATE STREAM RATINGS_WITH_CUSTOMER_DATA AS
SELECT * FROM RATINGS LEFT JOIN CUSTOMERS
ON R.ID=C.ID;
```

```
{
 "rating_id": 5313,
 "user_id": 3,
 "stars": 4,
 "route_id": 6975,
 "rating_time": 1519304105213,
 "channel": "web",
 "message": "worst. flight. ever. #neveragain"
}
```


Producer API



Kafka Connect



Join each rating to customer data

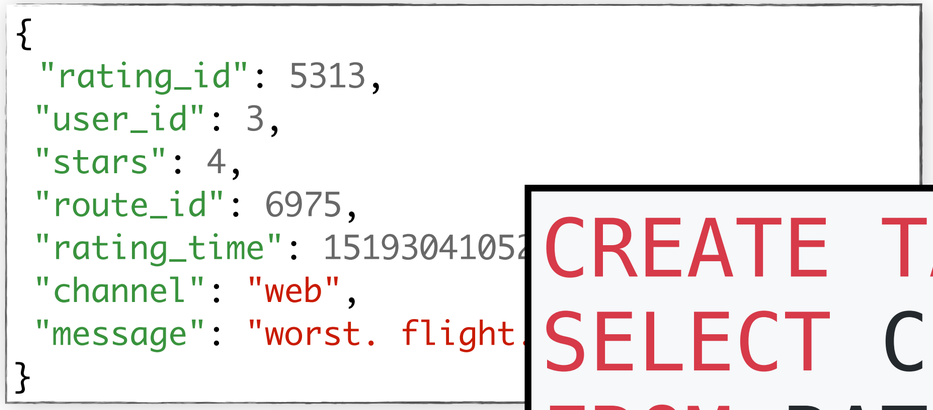
RATINGS_WITH_CUSTOMER_DATA

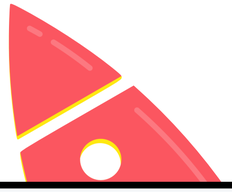
Filter for just PLATINUM customers

UNHAPPY_PLATINUM_CUSTOMERS

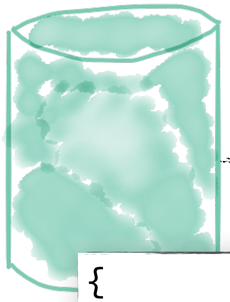

```
{
  "id": 3,
  "first_name
  "last_name"
  "email": "m
  "gender": "
  "club_statu
  "comments":
}
```


```
CREATE STREAM UNHAPPY_PLATINUM_CUSTOMERS AS
SELECT * FROM RATINGS_WITH_CUSTOMER_DATA
WHERE   STARS < 3
```

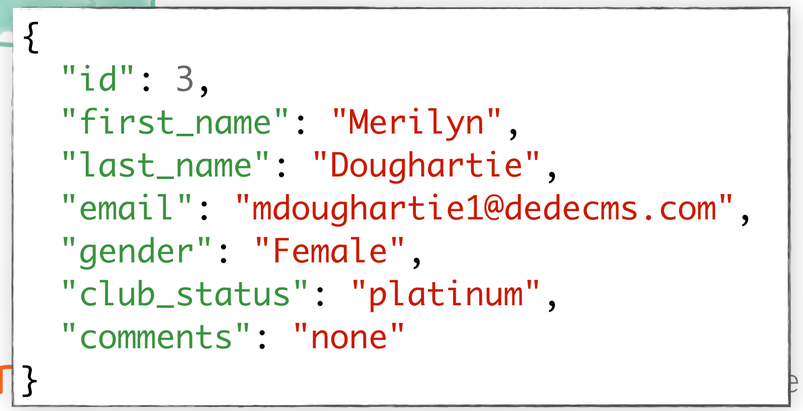




```
CREATE TABLE RATINGS_BY_CLUB_STATUS AS
SELECT CLUB_STATUS, COUNT(*)
FROM RATINGS_WITH_CUSTOMER_DATA
     WINDOW TUMBLING (SIZE 1 MINUTES)
GROUP BY CLUB_STATUS;
```



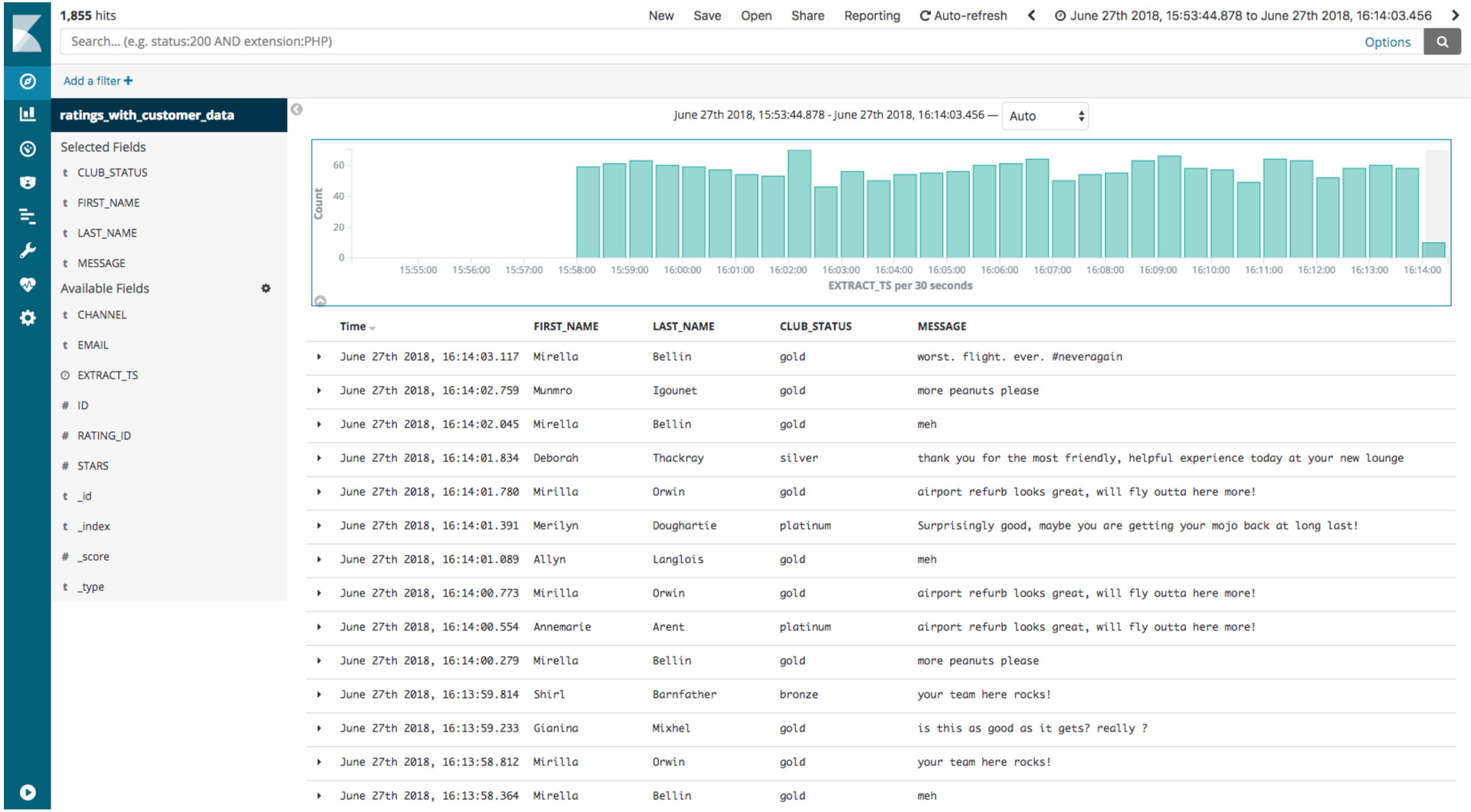
Kafka Connect



Aggregate per-minute by CLUB_STATUS

RATINGS_BY_CLUB_STATUS_1MIN

Stream to Elasticsearch



Time	FIRST_NAME	LAST_NAME	CLUB_STATUS	MESSAGE
June 27th 2018, 16:14:03.117	Mirella	Bellin	gold	worst. flight. ever. #neveragain
June 27th 2018, 16:14:02.759	Munmro	Igounet	gold	more peanuts please
June 27th 2018, 16:14:02.045	Mirella	Bellin	gold	meh
June 27th 2018, 16:14:01.834	Deborah	Thackray	silver	thank you for the most friendly, helpful experience today at your new lounge
June 27th 2018, 16:14:01.780	Mirilla	Orwin	gold	airport refurb looks great, will fly outta here more!
June 27th 2018, 16:14:01.391	Merilyn	Doughartie	platinum	Surprisingly good, maybe you are getting your mojo back at long last!
June 27th 2018, 16:14:01.089	Allyn	Langlois	gold	meh
June 27th 2018, 16:14:00.773	Mirilla	Orwin	gold	airport refurb looks great, will fly outta here more!
June 27th 2018, 16:14:00.554	Annemarie	Arent	platinum	airport refurb looks great, will fly outta here more!
June 27th 2018, 16:14:00.279	Mirella	Bellin	gold	more peanuts please
June 27th 2018, 16:13:59.814	Shirl	Barnfather	bronze	your team here rocks!
June 27th 2018, 16:13:59.233	Gianina	Mixhel	gold	is this as good as it gets? really ?
June 27th 2018, 16:13:58.812	Mirilla	Orwin	gold	your team here rocks!
June 27th 2018, 16:13:58.364	Mirella	Bellin	gold	meh

https://cnfl.io/ksql-workshop

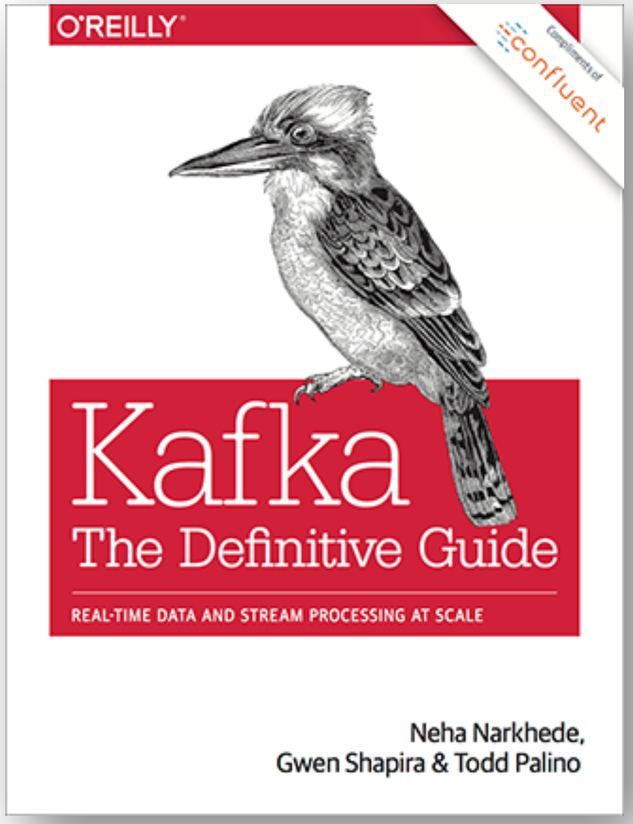
9. Joining Data in KSQL

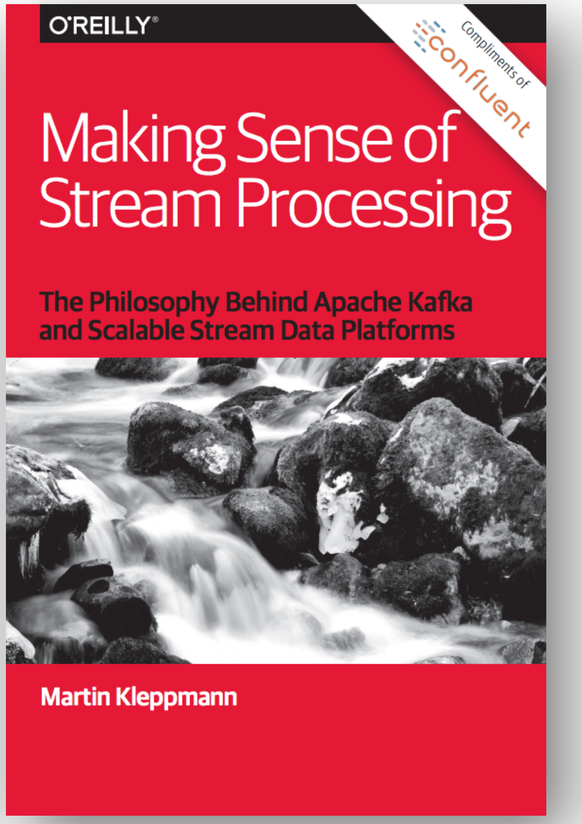
10. Daisy-chaining derived streams

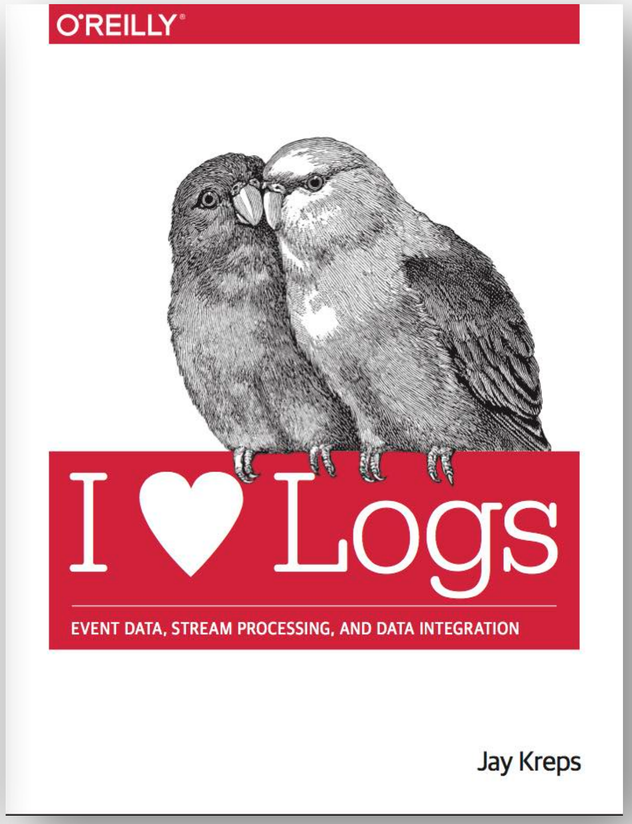
11. Streaming Aggregates

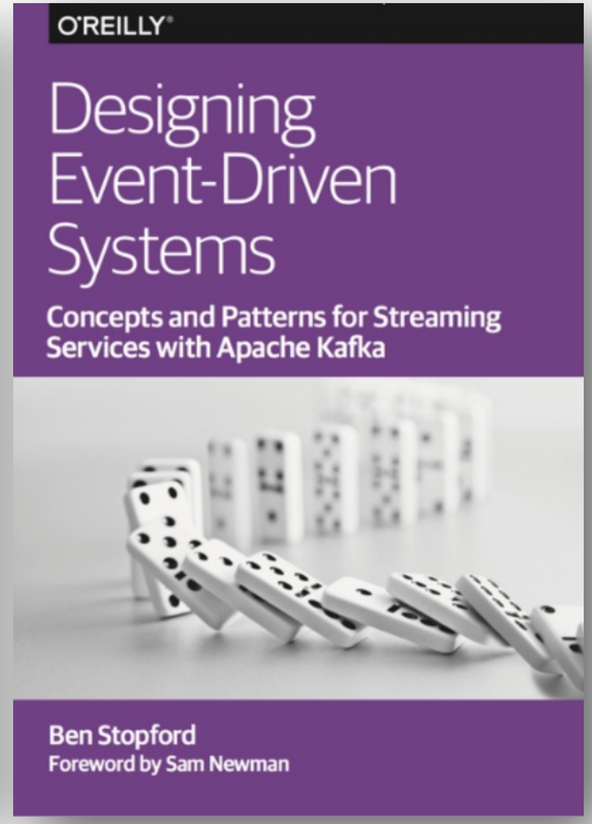
Free Books!

https://www.confluent.io/apache-kafka-stream-processing-book-bundle









https://www.confluent.io/ksql

http://cnfl.io/slack

@rmoff
robin@confluent.io

Useful links

- Embrace the Anarchy : Apache Kafka's Role in Modern Data Architectures Recording & Slides
- Look Ma, no Code! Building Streaming Data Pipelines with Apache Kafka and KSQL
- Steps to Building a Streaming ETL Pipeline with Apache Kafka and KSQL Recording & Slides
- https://www.confluent.io/blog/ksql-in-action-real-time-streaming-etl-from-oracle-transactional-data
- https://github.com/confluentinc/ksql/

#EOF